

Descriptive Data in the EDIT Platform for Cybertaxonomy

Maxime Venin, Agnes Kirchhoff, H  l  ne Fradin, Anton G  ntsches, Niels Hoffmann, Andreas Kohlbecker, Elise Kuntzelmann,   na Maiocco, Andreas M  ller, R  gine Vignes Lebbe, Walter G. Berendsohn

Abstract — This paper describes the integration of structured descriptive data in the EDIT platform for Cybertaxonomy. The platform is composed of several software modules supporting the taxonomic workflow from data capture and storage to publication. Descriptive data play an important role within the taxonomic work process. The integration of these data via import/export modules to and from the platform and the publication as natural language output or as keys are explained.

Index Terms —platform, software, taxonomy, description, key, natural language, structured descriptive data, SDD, Common Data Model, EDIT.



1 INTRODUCTION

One of the achievements of the European Distributed Institute of Taxonomy (EDIT) [1] is the Internet Platform for Cybertaxonomy, which provides software tools supporting and accelerating the taxonomic workflow (Fig. 1). “A main goal of the Platform is to provide an open architecture to allow connection and integration of existing applications and to provide new developments where necessary” [2]. The Platform is based on the Common Data Model (CDM), which is essentially a description of all data that can be used and edited in the Platform, such as taxon names and concepts, literature references, specimens, distributions, and structured and unstructured descriptive data. All data are stored in a repository known as the CDM Community Store. Different communities can set up their own Store, e.g. to work on a specific monograph, checklist or Flora/Fauna treatment.

The various Platform components are linked by interfaces to the Community Store, for example the Taxonomic Editor (EDITor) for data entry and the EDIT

M. Venin, H. Fradin, E. Kuntzelmann,   . Maiocco, and R. Vignes Lebbe are with the Mus  um National d'Histoire Naturelle (UPMC-MNHN), CP48, 57 rue Cuvier, 75231 Paris Cedex 05, France, E-mail: venin@mnhn.fr.

A. Kirchhoff, A. G  ntsches, N. Hoffmann, A. Kohlbecker, A. M  ller, W. G. Berendsohn are with the Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universit  t Berlin, K  nigin Luise Str. 6-8, 14195 Berlin, Germany, Email: a.kirchhoff@bgbm.org.

Data Portal for data publication (see Berendsohn, this volume).

The CDM code Library forms the heart of the Platform software. It enables the individual Platform components to interact. Software developers can use the Library to implement taxonomic software without having to re-create the functionality already developed.

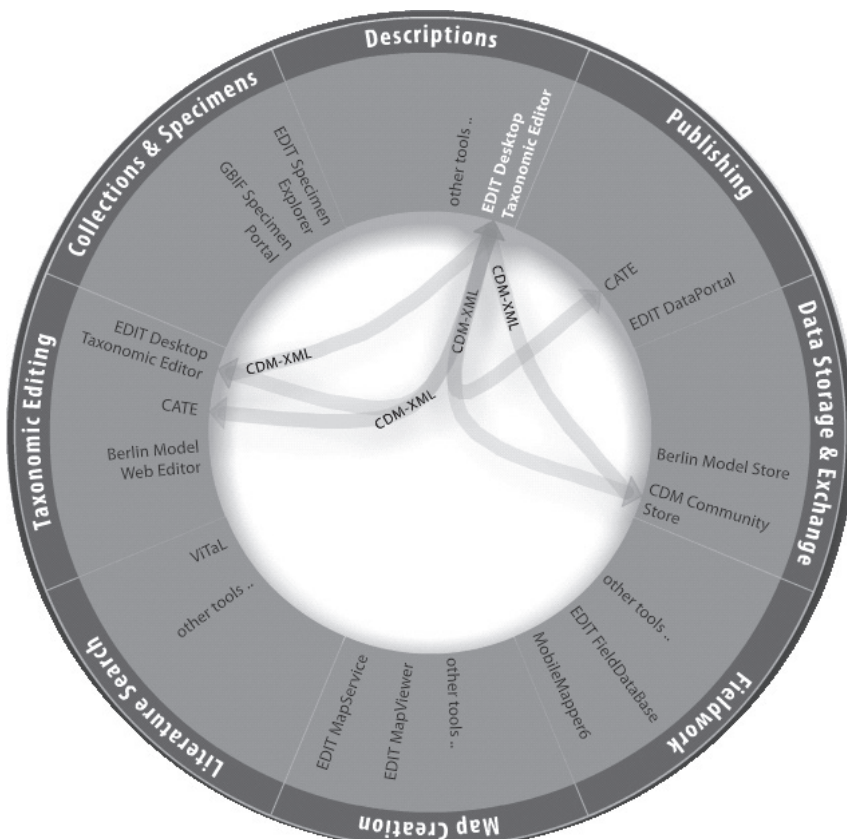


Fig. 1 – Overview over the software modules and functions of the EDIT Platform for Cybertaxonomy (EDIT Cybergate).

The EDIT Platform tools are designed to assist the taxonomist from fieldwork to publication of results, including the management of descriptive data, which play a key role in the taxonomic revision process.

Descriptive data are one of the most important categories of information produced by taxonomists when describing new species or performing taxonomic revisions. Traditionally, taxonomic descriptions were handled as text. However, storing and handling of descriptive data in a highly structured form has strong advantages: data exchange and integration is facilitated, and identification keys (both for printed output and interactive) as well as “natural language descriptions” can be generated automatically and in multiple languages.

There are several established software tools to manage and analyse descriptive data, some of them already existing for decades (e.g. DELTA [3]). Consequently,

it was decided at the outset of the EDIT project not to develop another application but to integrate existing descriptive tools into the EDIT Platform. The key for this is that the CDM complies with the SDD (Structured Descriptive Data) standard [4]. SDD is the current TDWG (Biodiversity Information Standards) standard for descriptive data. Many of the existing descriptive data managing tools (e.g. Lucid [5], Xper² [6], and DiversityDescriptions [7]) do support import and export of SDD conformant data, allowing their users to exchange descriptive data.

2 STRUCTURED DESCRIPTIVE DATA CONNECTED TO THE EDIT PLATFORM

2.1 THE SDD-CDM IMPORT AND EXPORT MODULE

An SDD-CDM import/export module was developed to integrate descriptive data with the EDIT Platform. Once in the CDM Community Store, the data can be published together with the other information on a specific taxon on the Internet, e.g. by means of the EDIT Data Portal software.

The SDD export from the CDM provides the possibility to use the different software tools and benefit from their specific analytical or output capabilities (interactive identification, comparison of taxa, statistics, etc.) (Fig. 2).

From the technical point of view the general idea behind the import of SDD elements is to create corresponding elements in the CDM in order to allow seamless export/import roundtrips.

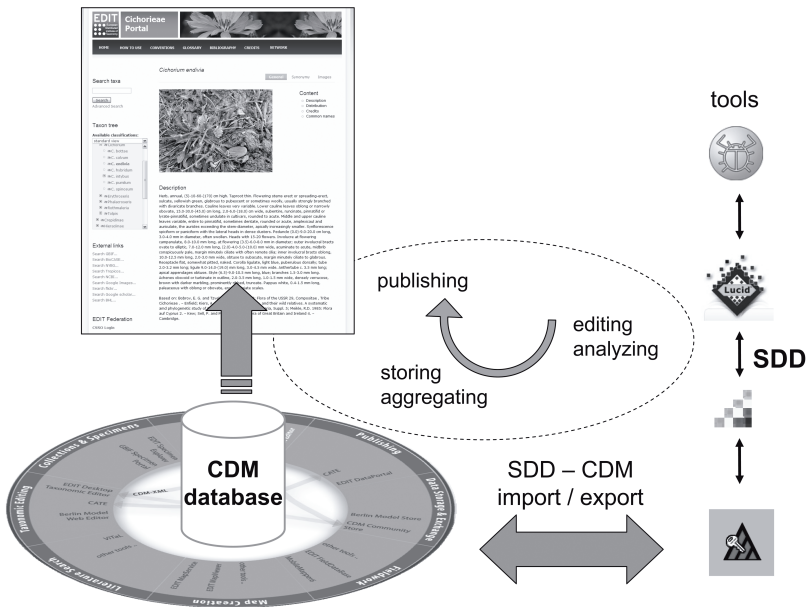


Fig. 2 – Data exchange between descriptive software tools and the EDIT platform for Cybertaxonomy.

2.2 DISPLAY OF DESCRIPTIVE DATA AS NATURAL LANGUAGE DESCRIPTIONS

The CDM Code Library now includes a feature to generate clear and easy to read output of the descriptive CDM data. The structure of the output can be pre-defined, which allows the scientists to keep a constant scheme, a very helpful feature when preparing output that has to adhere to a defined editorial standard. The natural language description output can be used for publications on the web or for print publications, or simply as a readable preview to control the content of the database.

2.3 DISPLAY OF KEYS

A next step in the processing of structured descriptive data is the possibility to automatically generate identification keys from the CDM data. The CDM Code Library now supports output in the form of dichotomous or polytomous identification keys that can be shown on taxon pages of higher taxa in the CDM data portals. Clickable links lead to other key entries or the identified taxa.

The integration of interactive keys will not be possible within the current EDIT project period, for this Platform users have to resort to existing tools. An example of such an integration is given by the CATE project [8], where the LUCID Player is used to provide interactive key functionality on the website.

3 FUTURE DEVELOPMENTS

The EDIT Taxonomic Editor (EDITor) is the main data entry tool of the EDIT Platform. It allows the editing and presentation of taxonomic information such as classifications, synonyms, taxonomic concepts, descriptions, distributions, specimens and literature references. As any other data in the EDIT Platform this kind of information is stored in the CDM Community Store.

As mentioned above, the EDIT Platform allows choosing among several software tools for the management of descriptive data. One of those is Xper², “a management system for storage, editing, analysis and online distribution of descriptive data” [9], which also dynamically creates interactive keys for identifying specimens. This software was chosen as a way forward for the integration of descriptive data into the EDIT platform, because it is Java-based, non-commercial, has been created by an EDIT partner, and it can be integrated with the EDIT Taxonomic Editor.

In the long term, full integration of Xper² with the Taxonomic Editor is the aim. A shorter term solution will be to enable Xper² to directly work with the data in a CDM Community Store. Xper² could then be opened via the EDITor, running as a separate application, but using the same data.

4 CONCLUSION

With respect to structured descriptive data, the current state of software development for the EDIT Platform for Cybertaxonomy can be summarised as follows:

With the SDD-CDM import/export module the integration of descriptive data into the Common Data Model has been completed.

The natural language module in the CDM library allows users to easily and rapidly generate output describing taxa and specimens. Thanks to the integration with other CDM objects and functions in the CDM Code Library, developers have a very broad range of possibilities to provide users with functions to create, use and publish natural language descriptions.

Generating simple keys is possible with the CDM library. It is an entirely automatic process based on the CDM Community Store. Once the descriptive data have been imported, a taxonomist can directly use this functionality without any extra work.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the support of: the EU 6th Network of Excellence Project EDIT (European Distributed Institute of Taxonomy, contract No 018340 - GOCE).

REFERENCES

- [1] N. N., "EDIT - European Distributed Institute of Taxonomy", <http://www.e-taxonomy.eu>, 2010.
- [2] P. Ciardelli, P. Kelbert, A. Kohlbecker, N. Hoffmann, A. Güntsch and W. G. Berendsohn, "The EDIT Platform for Cybertaxonomy and the taxonomic workflow: selected Components", *Lecture Notes in Informatics (LNI)*, vol. 154, pp. 625-638, 2009.
- [3] M. J. Dallwitz, "A flexible computer program for generating identification keys", *Syst. Zool.*, vol. 23, pp. 50-57, 1974.
- [4] G. Hagedorn et al., "The Structured Descriptive Data (SDD) w3c-xml-schema, version 1.1.", *TDWG*, <http://wiki.tdwg.org/wiki/bin/view/SDD/Version1dot1>, 2006.
- [5] N. N., "Lucidcentral", <http://www.lucidcentral.org/>. *Centre for Biological Information Technology, The University of Queensland, Brisbane*, 2010.
- [6] N. N., "Xper2", <http://lis-upmc.snv.jussieu.fr/lis/?q=en/resources/software/xper2>. *Laboratoire Informatique & Systématique, Paris*, 2010.
- [7] N. N., "DiversityDescriptions", <http://www.diversityworkbench.net/Portal/DiversityDescriptions>, 2008.
- [8] N. N., "The CATE Project", <http://www.cate-project.org/>, 2010.
- [9] V. Ung, G. Dubus, R. Zaragüeta-Bagils and R. Vignes Lebbe, "Xper2: introducing e-Taxonomy", *Bioinformatics*, vol. 26, no. 5, pp. 703-704, available at <http://bioinformatics.oxfordjournals.org/cgi/reprint/btp715v1.pdf>, Jan. 2010.